**Popularity Prediction**

***Problem presentation***

This software is supposed to offer o solution for tracking real live data from twitter platform. Our goal was to implement a project which can monitor a persons feed and based on his/her feed to create a series of statistics for that person. The purpose of this was to determine how some post are more popular from another. After we track down a post and make some basic statistics based on some features ( number of likes, number of re-tweets, number of comments ) we used that plain text to determine what type of feeling it has ( hateful , mean, good, bad, sad, happy etc. ).

Using this two information we can determine what kind of post are more popular among people , good post or bad posts. We think this software can illustrate a series of statistics very interesting because Twitter platform is really power full when we are talking about number of users that are using this platform. We tested our product on some famous persons like Donald J. Trump, Eminem, C. Ronaldo for having concrete data to display.

***State of the art***

Gathering this amount data we are going to create some early statistics ( using data from the API ) and stored them for creating a database for a neuronal network . This network will predict using the current data how legit is the person analyzing current tweets in compare with what is going to post next. All the results will be transformed in graphics ( one will be for current data, which will be shown instantly after we get the data from Twitter and the other will be what our classifier will decide). This classifier has the role of telling us what kind of sentiment a twitter has. We found this classifier on a site of a completion which was last year of building an app for determine a tweet sentiment using neuronal network. Instead of training a neuronal network which would be hard because we are limited on how many request we can made to twitter API we created a classifier.

Apps main component and some links explaining how we are going to implement it :

* **Collecting Data :** this component will have the goal to collect the data from Twitter API and prepare it to be stored efficient. ( <https://developer.twitter.com/en/docs/tweets/post-and-engage/overview> )
* **Parse Data and store it :** this component will have to use the data from “Collecting Data component” and start to parse it ( we still researching about what we are going to use from what twitter is providing us ) and store it ( maybe a database , or some static files ) link : https://jsonlint.com/
* **Drawing Graphics :** using the current stored data we are going to present to the user some nice statistics using multiple graphics, using a really nice library for drawing graphics in Python ( <https://matplotlib.org/> )
* **Main Interface :** this module will have the purpose to combine all the components from the project to display a nice interface for user interaction
* **AI Module :** this component will work offline after the app has enough data to train a neuronal network for getting some predictions using current database. After the module can work will have his specific graphic showing the results.
* **Statistics Module:** this module will calculate and determine using the current data from Twitter API some preliminary statistics of the current profile. This results will be send to “Graphics component" to display them nice.

***Solution***

We created a software which we first registered at Twitter API to gain access to their data. Unfortunately because this is only a beta version of this project we didn`t pay at Twitter to have full access to request their API so the amount of data is pretty limited. This kind of limitation forced us to store some data we gathered to have a larger database to accurate our statistics. We can made per day only 250 request at Twitter which means only we can make an evolution for a specific person per day. We track down a post at an interval of 60 seconds because Twitter post of influent persons grown exponentially in the first hour , maybe two and than it stops.

Tracking down post in this way provide us the chance to follow up how fast a twitter get his like. E.g. for Donald Trump if he posted something, in max 1 minute he has 20k+ likes which is really huge. For storing this data we created some CSV file which are pretty easy to parse in python.

For the second part of the project which means determining the feeling of the tweet, we find on internet some database used for training a neuronal network. We parsed the data from tweeter and send it to our classifier. The classifier was built based on the data parsed from the sentimental database. Using this classifier and our data we can calculate with a good precision what kind of sentiment a tweet has.

***Other Solutions***

On our way of finding what kind of project we wanted to built we inspired from a paper of analyzing people influence and this was their solution for twitter influence.

C. TwitterRank

TwitterRank measures the influence taking both the topical similarity between users and the link structure into account. In a dataset prepared for this study, it is observed that 1)72.4% of the users follow more than 80% of their followers, and (2) 80.5% of the user have 80% of their friends follow them back.[4]Our study reveals that the presence of “reciprocity” can be explained by phenomenon of homophily.Based on this finding, TwitterRank, an extension of PageRank algorithm, is proposed to measure the influence of users in Twitter.

TwitterRank measures the influence taking both the topical similarity between users and the link structure into account. Experimental results show that TwiterRank outperforms the one Twitter currently uses and other related algorithms, including the original PageRank and Topic-sensitive PageRank. First, it potentially brings order to the real-time web in that it allows the search results to be sorted by the authority/influence of the contributing twitterers giving a timely update of the thoughts of influential twitterers. Second, Twitter is also a marketing platform. Targeting those influential users will increase the efficiency of the marketing campaign. For example, a handphone manufacturer can engage those twitterers influential in topics about IT gadgets to potentially influence more people .

There are also applications that utilize Twitter to gather opinions and information on particular topics. Identifying influential twitterers for interesting topics can improve the quality of opinions gathered. PageRank improves over in-degree by considering the link structure of the whole network. Nevertheless, Pagerank ignores the interests of twitterers, which affects the way twitterers influence one another. Our proposed approach addresses the shortcomings of in-degree and PageRank by taking into account both the link structure and topical similarity among twitterers.

In the context of Twitter, homophily implies that a twitterer follows a friend because she is interested in some topics the friend is publishing, and the friend follows back because she finds they share similar topical interest.

***Conclusions & Future Work***

In conclusion our project had a research purpose for studying humans on the social environment. Our result were pretty interesting because we find out that people were reacting much faster and more aggressive on hateful post, or mean post in comparison with good posts or science posts. This kind of results were pretty good to anticipate because we can watch on how our news posts are created and all of them try to show what is more mean, hard, bad, sensational etc. instead of informing you with useful information . In the end we created an Agent which collects data, store it and use it to create statistics and concrete results of tweets.

Using them we can continue our work for future applications. We think of creating a product which use this modules to tell a person how her/his tweets affects direct his popularity. We can let our Agent to monitor her/his activity for saying 1 month and then come with concrete data and graphics to show the customer how his/her activity on twitter influenced his/her follow list.

Tracking down some random users from customer followers list we can observe them for a while to see how our customer influenced his tweets or his/her activity. Would be really interesting to see how other persons are influenced by their idols or models in their day by day activities. A study says that we spent almost 3 hours per day on social media and would be really nice if we can create some statistics on other persons based on who is their model or idol ( in our case our customer ) .

***Bibliography*** :

* <https://developer.twitter.com/en/docs/tutorials.html>
* <http://socialmedia-class.org/twittertutorial.html>
* <https://www.w3resource.com/API/twitter-rest-api/>
* <https://matplotlib.org/tutorials/index.html>
* <https://www.digitalocean.com/community/tutorials/how-to-build-a-machine-learning-classifier-in-python-with-scikit-learn>
* <https://pdfs.semanticscholar.org/1264/04e06c8ecc4a73f80827f2a2fcda1d6066b4.pdf>
* <https://www.programmableweb.com/news/twitter-influence-api-makes-klout-billionaire/2011/10/04>
* <https://stackabuse.com/accessing-the-twitter-api-with-python/>
* <https://www.tutorialspoint.com/python/python_matplotlib.htm>
* <https://www.kdnuggets.com/2018/08/emotion-sentiment-analysis-practitioners-guide-nlp-5.html>
* <https://datascience.stackexchange.com/questions/11220/training-dataset-for-sentiment-analysis-of-movie-reviews>